501.43144X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s):	MURAKAMI, Toshihiko
Serial No.:	Not yet assigned
Filed:	September 17, 2003
Title:	DATA TRANSFER METHOD
Group:	Not yet assigned

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

September 17, 2003

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on Japanese Patent Application No.(s) 2003-038097, filed February 17, 2003.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP

Carl I. Brundidge
Registration No. 29,621

CIB/alb
Attachment
(703) 312-6600

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: ２００３年　２月１７日

出願番号
Application Number: 特願２００３－０３８０９７

[ST. 10/C]: ［ＪＰ２００３－０３８０９７］

出願人
Applicant(s): 株式会社日立製作所

２００３年　７月３１日

特許庁長官
Commissioner,
Japan Patent Office 今井康夫

出証番号　出証特２００３－３０６１１５２

【書類名】　特許願
【整理番号】　K201032I
【提出日】　平成15年 2月17日
【あて先】　特許庁長官　殿
【国際特許分類】　G06F 12/02
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内
　【氏名】　村上　俊彦
【特許出願人】
　【識別番号】　000005108
　【氏名又は名称】　株式会社日立製作所
【代理人】
　【識別番号】　100099298
　【弁理士】
　【氏名又は名称】　伊藤　修
　【連絡先】　０３－３２５１－３８２４
【選任した代理人】
　【識別番号】　100099302
　【弁理士】
　【氏名又は名称】　笹岡　茂
【手数料の表示】
　【予納台帳番号】　018647
　【納付金額】　21,000円
【提出物件の目録】
　【物件名】　明細書　1
　【物件名】　図面　1
　【物件名】　要約書　1
【プルーフの要否】　要

【書類名】　　明細書

【発明の名称】　データ移行方法

【特許請求の範囲】

【請求項１】　複数の計算機と複数の記憶装置と、前記計算機と記憶装置を相互に接続する中継装置と、前記計算機と記憶装置と中継装置を管理する管理装置とで構成された計算機システムにおけるデータ移行方法であって、

前記管理装置は、前記複数の計算機に対して前記記憶装置の仮想的な記憶領域を設定し、該設定の内容の情報を第１の情報として保持し、

前記中継装置は、前記第１の情報を元に作成された第２の情報を保持し、

前記仮想的な記憶領域は前記それぞれの記憶装置内の記憶領域あるいは複数の記憶装置内の記憶領域が結合されてなる記憶領域に対応し、

前記中継装置は、前記第２の情報から一つの仮想的な記憶領域を選択し、該選択した仮想的な記憶領域が、複数の記憶装置内の記憶領域が結合されてなる記憶領域である場合を契機として、該複数の記憶装置間でのデータの移行を行い、該結合にかかわる記憶装置を少なくすることを特徴とするデータ移行方法。

【請求項２】　請求項１記載のデータ移行方法において、

前記中継装置は、前記第２の情報を参照し、前記仮想的な記憶領域に対応していない比較的小さな容量の記憶領域が増加してきた場合を契機として、該複数の記憶装置間でのデータの移行を行い、前記仮想的な記憶領域に対応していない比較的小さな容量の記憶領域を少なくすることを特徴とするデータ移行方法。

【請求項３】　請求項１または請求項２記載のデータ移行方法において、

前記第２の情報は前記第１の情報を元に更新されることを特徴とするデータ移行方法。

【請求項４】　請求項１または請求項２記載のデータ移行方法において、

前記複数の記憶装置間でデータ移行をする際に、データ移行の単位ごとにデータ移行未完かデータ移行済かを示す第３の情報を設定し、該データ移行の単位がデータ移行されるごとに前記データ移行未完をデータ移行済に変更し、データ移行の進行状況を更新することを特徴とするデータ移行方法。

【請求項５】　請求項３記載のデータ移行方法において、

前記第二の情報は前記仮想的な記憶領域毎にデータ移行中であるかどうかの状態を示すフラグを有し、

前記中継装置は、前記計算機が前記仮想的な記憶領域をアクセスするときに、前記第二の情報の該当する前記仮想的な記憶領域の前記フラグがデータ移行中である場合は、前記第三の情報によりデータ移行元またはデータ移行先のどちらにアクセスを行うべきかを判断し、データ移行中に前記計算機から前記仮想的な記憶領域へのデータのアクセスを中断させないことを特徴とするデータ移行方法。

【請求項６】　請求項１乃至請求項５のいずれかの請求項記載のデータ移行方法において、

前記中継装置は、前記仮想的な記憶領域の一部であるデータ移行を行う記憶領域のデータを、データ移行先の記憶領域へ直接コピーすることを特徴とするデータ移行方法。

【請求項７】　請求項１乃至請求項５のいずれかの請求項記載のデータ移行方法において、

前記中継装置は、前記仮想的な記憶領域の一部であるデータ移行を行う記憶領域のデータを、データ移行のために予め用意している前記記憶装置内の記憶領域に一旦コピーした後に、データ移行先の記憶領域へ間接的にコピーすることを特徴とするデータ移行方法。

【請求項８】　請求項１乃至請求項５のいずれかの請求項記載のデータ移行方法において、

前記中継装置は、前記仮想的な記憶領域の一部であるデータ移行を行う記憶領域のデータを、前記複数の記憶装置の記憶領域の未使用の記憶領域に一旦コピーした後に、データ移行先の記憶領域へ間接的にコピーすることを特徴とするデータ移行方法。

【請求項９】　請求項１乃至請求項５のいずれかの請求項記載のデータ移行方法において、

前記中継装置は、データ移行時の前記仮想的な記憶領域の一部であるデータをデータ移行後も、データ移行前の仮想的な記憶領域に一時的にまたは指定した期間残しておくことにより、データ移行後の新しい仮想的な記憶領域のバックアッ

プデータとすることを特徴とするデータ移行方法。

【請求項１０】　請求項１乃至請求項５のいずれかの請求項記載のデータ移行方法において、

前記中継装置は、前記仮想的な記憶領域の一部であるデータ移行を行う記憶領域のデータを、データ移行のために予め用意している前記中継装置内の記憶領域に一旦コピーした後に、データ移行先の記憶領域へ間接的にコピーすることを特徴とするデータ移行方法。

【請求項１１】　請求項１記載のデータ移行方法において、

前記計算機システム内の前記中継装置を冗長構成とした場合、前記管理装置は、前記第一の情報を全ての前記中継装置へと配布し、前記第二の情報の情報源とすることを特徴とするデータ移行方法。

【請求項１２】　請求項１記載のデータ移行方法において、

前記計算機システム内の前記中継装置の内部の構成要素を冗長構成とし、前記第二の情報を持つ構成要素が冗長構成となる場合、当該冗長構成の構成要素間で第二の情報の同期を常にとることにより、当該冗長構成の構成要素の一つが障害となった場合に、その他の当該冗長構成の構成要素の前記第二の情報を前記中継装置が利用することを特徴とするデータ移行方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、複数の記憶装置の記憶領域を仮想化して利用する仮想化技術に関わり、記憶装置間または記憶装置内のデータ移行方法に関する。

【０００２】

【従来の技術】

複数の計算機と複数の記憶装置をネットワークで接続するＳＡＮ（Storage Area Network）では、記憶装置の記憶領域を仮想化して利用させる技術がある（例えば、非特許文献１参照）。

計算機と記憶装置の間のどの位置で仮想化を行うかによりいくつかの方法がある。

第一はサーバアプリケーションが実行される計算機で、ストレージ管理ソフトウェアやボリューム管理ソフトウェアにより仮想化を行う方法である。

第二は記憶装置の手前に記憶装置を接続するインタフェースを複数用意した計算機を置き、この計算機で仮想化を行う方法や記憶装置そのものが仮想化を行う方法である。

第三はＳＡＮを構成するネットワーク機器や管理サーバ装置で仮想化を行う方法で、制御用とデータ用のＩ／Ｏ（Input／Output）を同じネットワークで扱うインバンド方法と、制御用のＩ／Ｏを流すネットワークを別途設けてデータ用のＩ／Ｏとは別に扱うアウトオブバンド方法の二つの方法がある。

【非特許文献１】

Evaluator Group社の技術白書である「Virtualization of Disk Storage（2000年9月）」

【０００３】

【発明が解決しようとする課題】

従来の技術で示した第二または第三の仮想化の方法で、ファイバチャネルスイッチやイーサネット（登録商標）でｉＳＣＳＩ（Internet Small Computer System Interface）を使用してＳＡＮを構成する場合のＬＡＮスイッチのような中継装置や、ファイバチャネルやイーサネット（登録商標）のＨＢＡ（Host Bus Adaptor）を複数備えるサーバベースの装置が仮想化された記憶領域（仮想ボリューム）を提供している場合に、仮想ボリュームを複数の記憶装置の記憶領域の結合で構成していると、仮想ボリュームのアクセスの状況によっては、複数の記憶装置の中の、例えば、第一の記憶装置、第二の記憶装置、第三の記憶装置という具合に、次々にアクセスする宛先の記憶装置が変化し、中継処理負荷が増加するという問題が発生する。

本発明の第一の目的は、比較的少ない記憶装置からの記憶領域で仮想ボリュームを構成するように、運用中にデータ移行を行うことにより、中継処理負荷を低減させる方法を提供することにある。

また、複数の記憶装置の記憶領域から仮想ボリュームを構成し、領域の拡大または縮小を随時行っていくと、比較的小さな容量の未使用の記憶領域が増加する

場合が考えられ、多数の小さな容量の記憶領域で仮想ボリュームを構成すると、中継装置における仮想ボリュームと実際の記憶領域との対応の変換処理の処理負荷が増加する。

そして、比較的大きな仮想ボリュームを比較的少ない数の未使用の記憶領域から構成することができなくなるという問題も発生する。

本発明の第二の目的は、比較的小さな容量の未使用の記憶領域が少なくなるように、既に仮想ボリュームを構成している一つ以上の記憶領域を別の記憶領域へと、運用中にデータ移行を行うことにより、この問題を低減させる方法を提供することにある。

【０００４】

【課題を解決するための手段】

上記目的を達成するため、本発明は、複数の計算機と複数の記憶装置と、前記計算機と記憶装置を相互に接続する中継装置と、前記計算機と記憶装置と中継装置を管理する管理装置とで構成された計算機システムにおけるデータ移行方法であり、

管理装置は、複数の計算機に対して記憶装置の仮想的な記憶領域を設定し、該設定の内容の情報を第１の情報として保持し、中継装置は、第１の情報を元に作成された第２の情報を保持し、

仮想的な記憶領域は前記それぞれの記憶装置内の記憶領域あるいは複数の記憶装置内の記憶領域が結合されてなる記憶領域に対応し、

中継装置は、第２の情報から一つの仮想的な記憶領域を選択し、該選択した仮想的な記憶領域が、複数の記憶装置内の記憶領域が結合されてなる記憶領域である場合を契機として、該複数の記憶装置間でのデータの移行を行い、該結合にかかわる記憶装置を少なくする。

また、中継装置は、第２の情報を参照し、仮想的な記憶領域に対応していない比較的小さな容量の記憶領域が増加してきた場合を契機として、複数の記憶装置間でのデータの移行を行い、前記仮想的な記憶領域に対応していない比較的小さな容量の記憶領域を少なくする。

【０００５】

【発明の実施の形態】

本発明の第一の実施形態を図１から図１３を参照して説明する。

第一の実施形態では、比較的少ない記憶装置からの記憶領域により仮想ボリュームの構成を行うように、運用中にデータ移行を行う場合について説明する。

図１は、本発明における計算機システムの構成例を示す図である。

サーバ装置１７０がスイッチ１００を介してストレージ装置１８０と接続される。

スイッチ１００は、ファイバチャネルスイッチやＬＡＮスイッチであるが、本発明の実施例の説明の中継装置としては、イーサネット（登録商標）スイッチの場合を例として示す。

スイッチ１００は、サーバ装置１７０またはストレージ装置１８０等を接続するための物理層とデータリンク層の制御を行う複数の接続インタフェース１１０（図１では、Ether MACと表記している）と、送受信するパケットのレイヤ２から上位（図中はレイヤ７までと示している）のヘッダまたは情報部分の情報を元に宛先となる接続インタフェースの決定や必要時にはパケットの内容の変換を行う複数のルーティング制御部（Routing Control）１２０と、複数のルーティング制御部を接続するクロスバースイッチ（Crossbar Switch）１３０と、スイッチ１００の装置管理や経路制御プロトコルの計算を行うスイッチ管理部（Switch Manager）１４０と、クロスバースイッチ１３０とスイッチ管理部１４０を接続する内部バス等の内部通信線１０１Ｃで構成される。

【０００６】

ルーティング制御部１２０は、ＣＰＵ１２６、プログラムやテーブルを格納するメインメモリ（MEM）１２７、レイヤ２からレイヤ７のパケットのフォワーディング制御部（Layer2-7 Forwarding Engine）１２１、パケットのヘッダ検索部（Search Engine）１２２とそのテーブル１２５を格納するメモリ（MEM）１２３を有し、これらが内部バス等の内部通信線１２８で接続されている。

スイッチ管理部１４０は、ＣＰＵ１４１、プログラムや経路制御テーブル等のテーブル１４３を格納するメインメモリ（MEM）１４２、管理用のイーサネット（登録商標）（Ether）１４５、クロスバースイッチへの内部通信線１０１Ｃと

接続するための内部バス制御１４４を有し、これらが内部バス等の内部通信線１４６で接続されている。

スイッチ１００には、さらに管理用のコンソールまたはイーサネット（登録商標）のポート１６１を介して管理装置１６０や、接続インタフェース（Ether MAC）１１０を介してストレージ装置１８０の実ボリューム、またはスイッチ１００が提供する仮想ボリュームのバックアップ等の機能をスイッチと連携して行うための複数の連携サーバ１５０が接続される。

連携サーバ１５０は、ＣＰＵ１５１、プログラムやボリューム管理テーブル等のテーブル１５３を格納するメインメモリ（MEM）１５２、イーサネット（登録商標）（Ether）１５６、ハードディスク（hard disk）１５７を有し、これらが内部バス制御１５４を介して内部バス等の内部通信線１５５で接続されている。

【０００７】

図２は、スイッチ１００のスイッチ管理部１４０内のメインメモリ１４２にあるプログラムおよびテーブルの構成例を示す図である。

メインメモリ１４２のプログラム領域２０１は、オペレーティングシステム２０３、経路制御プロトコル２０４、実／仮想ボリューム管理２０７、装置管理２０９、スイッチ宛パケット処理２１１、データ移行処理２１２等から構成され、テーブル領域１４３は、経路制御プロトコル２０４と関係するルーティングテーブル２０５とフィルタリングテーブル２０６、実／仮想ボリューム管理２０７と関係するボリューム管理テーブル２０８、装置管理２０９と関係する装置管理テーブル２１０、ユニット使用状況管理テーブル２１３等から構成される。装置管理テーブルは接続されている装置および内蔵する装置の種類、構成情報、性能情報等を記録したテーブルであり、図４、５における装置管理テーブルも同様のものである。

【０００８】

図３は、スイッチ１００のルーティング制御部１２０内のメインメモリ１２７およびメモリ１２３（１２５）にあるプログラムおよびテーブルの構成例を示す図である。

メインメモリ１２７はオペレーティングシステム３０１、ハード処理不可パケ.

ット処理３０３、スイッチ管理部との通信処理３０２等から構成され、メモリ１２５は、スイッチ管理部との通信処理３０２と関係するルーティングテーブル３０４、フィルタリングテーブル３０５、ボリューム管理テーブル３０６等から構成される。

【０００９】

図４は、連携サーバ１５０のメインメモリ１５２にあるプログラムおよびテーブルの構成例を示す図である。

メインメモリ１５２のプログラム領域４０１は、オペレーティングシステム４０２、実／仮想ボリューム管理４０３、バックアップ機能処理４０４、装置管理４０６、スイッチまたは管理装置との通信処理４０８等から構成され、テーブル領域１５３は、実／仮想ボリューム管理４０３およびバックアップ機能処理４０４と関係するボリューム管理テーブル４０５、装置管理４０６と関係する装置管理テーブル４０７等から構成される。

【００１０】

図５は、管理装置１６０にあるプログラムおよびテーブルの構成例を示す図である。

メインメモリ５０１のプログラム領域５０２は、オペレーティングシステム５０４、ストレージ管理ソフト５０５として装置管理５０８や実／仮想ボリューム管理５０６、スイッチまたは連携サーバとの通信処理５１０等から構成され、テーブル領域５０３は、実／仮想ボリューム管理と関係するボリューム管理テーブル５０７、装置管理５０８と関係する装置管理テーブル５０９等から構成される。

【００１１】

図２から図５で示したボリューム管理テーブル２０８、３０６、４０５、５０７は、図６から図８に示すユニット管理テーブル６００、論理ユニット管理テーブル７０１、仮想ボリューム管理テーブル７０２から構成される。

図１で示した計算機システムの管理者が管理装置１６０により、ボリューム管理テーブル５０７を作成または更新し、スイッチ１００のスイッチ管理部１４０のボリューム管理テーブル２０８、連携サーバ１５０のボリューム管理テーブル

４０５へ配布し、スイッチ管理部１４０はルーティング制御部１２０のボリューム管理テーブル３０６へ配布する。

【００１２】

図６は、ユニット管理テーブルの構成例を示す図である。

ストレージ装置１８０は計算機１７０やスイッチ１００と接続するための複数の接続ポート６０１、複数のディスク６０４をアクセスするためのディスクアダプタ６０３を有し、複数の接続ポート６０１と複数のディスクアダプタ６０３がある場合は、内部スイッチ６０２で各接続ポートと各ディスクアダプタが接続されている。

ユニット管理テーブル６００は、複数のエントリ６０５から構成される。

ディスク毎のエントリ６０５は、ユニットＩＤ、ＭＡＣ（Media Access Control）アドレス、ＩＰアドレス、ポートＩＤ、ディスク番号、ブロック数（説明を簡単にするため、１メガバイトを１ブロックとしている）、容量等の情報を持つ。

【００１３】

図７は、論理ユニット管理テーブルと仮想ボリューム管理テーブルの構成例を示す図である。

論理ユニット管理テーブル７０１は、複数のエントリ７０１Ａから構成される。

論理ユニット毎のエントリ７０１Ａは、論理ユニット（ＬＵ）ＩＤ、図６で示したユニットＩＤ、開始ＬＢＡ（Logical Block Address）、終了ＬＢＡ、サイズ、状態等の情報を持つ。

仮想ボリューム管理テーブル７０２は、複数のエントリ７０３から構成される。

仮想ボリューム毎のエントリ７０３は、仮想ボリュームＩＤ、構成論理ユニット（ＬＵ）の組、サイズ等の情報と、移動可否を示す情報を持ち、エントリ７０３Ａでは移動可否は禁止となっており、エントリ７０３Ｂでは移動可否は可能となっている。

一例として示されている仮想ボリューム７０４は、仮想ボリュームテーブル７

０２のエントリ７０３Ａが示すように、論理ユニットＩＤがＶＬＵ００とＶＬＵ０１の論理ユニットの結合として構成されており、移動可否は禁止である。

【００１４】

図８は、図７と同様に論理ユニット管理テーブルと仮想ボリューム管理テーブルの構成例を示す図である。

仮想ボリューム８０４は、仮想ボリュームテーブル７０２のエントリ７０３Ｂが示すように、論理ユニットＩＤがＶＬＵ１０とＶＬＵ１１の論理ユニットの結合として構成されており、移動可否は可能である。

図７と図８の仮想ボリュームは、同じストレージ装置１８０内のディスク６０４で構成されているものが図７、別のストレージ装置１８０内のディスク６０４で構成されているものが図８であり、本発明のデータ移行方法は図８のような場合にデータ移行を行う。

図９は、図８の仮想ボリューム８０４でデータ移行を行った場合の処理の概要を説明するための図である。

【００１５】

図１０はスイッチ管理部１４０のデータ移行処理２１２がデータ移行を行う契機を得るための処理シーケンスの例を示す図である。

図１０の処理シーケンスにおいて、仮想ボリューム、例えば仮想ボリューム８０４を選択すると（ステップ１００１）、次に、選択した仮想ボリュームはデータ移動禁止であるか判定し（ステップ１００８）、判定の結果がデータ移動禁止であればデータ移行を行わず、データ移動禁止でなければステップ１００２に進む。

ステップ１００２で構成LUの調査をする。選択した仮想ボリューム８０４の場合には、仮想ボリューム管理テーブル７０２のエントリ７０３Ｂにより、構成されている論理ユニットはＶＬＵ１０とＶＬＵ１１の結合であることが判る。

ステップ１００３では、構成LUのIPアドレスが全一致であるかを判定し、全一致であればデータ移行を行わず、全一致でなければステップ１００４に進む。仮想ボリューム８０４の場合には、論理ユニット管理テーブル７０１のエントリ８０１Ａ１と８０１Ａ２によりユニットＩＤはそれぞれ、ＲＵＡ０１００とＲＵＢ

０００であり、ユニット管理テーブル６００によりＩＰアドレスが全一致しないので、ステップ１００４に進む。

ステップ１００４では構成LUを選択し、ステップ１００５に進む。ここでは仮想ボリューム８０４の場合、ＲＵＡ０１００の論理ユニットを選択したとする。

ステップ１００５では、選択した構成LUの記憶装置内に別の構成LUからデータ移行可能な実ユニット内の未使用領域が存在するか判定し、存在すればステップ１００７に進み、存在しなければステップ１００６に進む。仮想ボリューム８０４の場合、ＲＵＡ０１００の論理ユニットが選択されており、ステップ１００５の判定の結果、別の構成LUであるＲＵＢ００００からのデータ移行可能な領域が存在するのでステップ１００７に進む。

ステップ１００７ではデータ移行処理を行う。仮想ボリューム８０４の場合、ＲＵＢ００００内のＶＬＵ１１をＲＵＡ０１００へデータ移行する。

ステップ１００６では、別の構成LUがあればステップ１００４に戻り、無ければデータ移行は行わない。

【００１６】

図１０に示す処理シーケンスは、データ移行を行わない場合や（ステップ１００３のＹＥＳの場合）、行うことができない場合もある（ステップ１００６のＮＯの場合）。

また、図１０に示す処理シーケンスは、開始の指示を管理装置等からの管理者の指示により行う場合や、ストレージ管理ソフト５０５のスケジュールにより行う場合もある。

【００１７】

図１１はデータ移行処理１００７の処理シーケンスの例を示す図である。

図１２は仮想ボリューム内の論理ユニットの状態を管理するテーブル１２００の例を示す図である。図において状態１２０６はデータ移行に係らないときはアイドルとなる。

図１０で示したデータ移行を行う契機を得るための処理シーケンスにおいて、ユニットＲＵＡ０１００にデータ移行可能な移行先があるということで、論理ユニットＶＬＵ１１と同じサイズの論理ユニットＶＬＵ１２をエントリ１２０９の

ように設定する。その際に状態フィールド１２０６を移行先と設定しておく（ステップ１１０１）。

移行元の論理ユニットＶＬＵ１１は、エントリ１２０８に示すように状態フィールド１２０６を移行元にする（ステップ１１０２）。

ここまでの処理は図９の（１）ＬＵ作成である。

次に移行元の論理ユニットＶＬＵ１１から移行先の論理ユニットＶＬＵ１２にデータを移行し（ステップ１１０３）、仮想ボリュームを構成する論理ユニットを移行元の論理ユニットＶＬＵ１１から移行先の論理ユニットＶＬＵ１２へと変更する（ステップ１１０５）。

ここまでの処理は図９の（２）データ移行と（３）論理ユニット切替えである。

次に移行先の論理ユニットＶＬＵ１２の状態をエントリ１２１０に示すように状態フィールド１２０６をアイドルと設定し、また、仮想ボリュームＶＶＯＬ０２ｋ移行可否を禁止に設定し（ステップ１１０６）、移行元の論理ユニットＶＬＵ１１を削除し（ステップ１１０７）、移行元論理ユニットＶＬＵ１１内のデータを削除する（ステップ１１０８）。

ここまでの処理は図９の（４）論理ユニットＶＬＵ１１削除と（５）論理ユニットＶＬＵ１１内データ削除である。

なお、（４）の論理ユニットＶＬＵ１１削除と（５）の論理ユニットＶＬＵ１１内データ削除とをせずに、一時的に、または指定した期間だけ論理ユニットＶＬＵ１１と論理ユニットＶＬＵ１１内データをバックアップ用として残しておくようにしてもよい。

【００１８】

図１１のデータ移行処理（ステップ１１０３）は、データ移行の最中のＲｅａｄやＷｒｉｔｅのＩ／Ｏも処理する必要があるため、図１３に示すように、データ移行をブロック単位毎にデータ移行の未完を管理するブロックｂｉｔｍａｐテーブル１３０１を有する。

移行元の論理ユニットＶＬＵ１１から移行先の論理ユニットＶＬＵ１２にブロック単位でデータを移動し（ステップ１１０４Ａ）、ブロックｂｉｔｍａｐテー

ブル１３０１を更新し（ステップ１１０４Ｂ）、すべてのデータの移動が完了するまで同様の処理を繰り返す（ステップ１１０４Ｃ）。

データ移行の最中のＩ／Ｏ処理（ステップ１１０９）は、図１３に示すような処理マトリックス１３０２のような処理を行う。

非データ移行領域に対するＩ／Ｏ処理１３０３は、該当する記憶領域へそのまま適用され、データ移行領域に対するＩ／Ｏ処理１３０４または１３０５は、ブロックｂｉｔｍａｐテーブル１３０１の状態によりデータ移行元またはデータ移行先の論理ユニットの記憶領域へ適用される。

【００１９】

以上の図１０から図１３で示したスイッチ管理部１４０のデータ移行処理２１２により、ボリューム管理テーブル２０８の構成に変更があった場合は、ルーティング制御部１２０のボリューム管理テーブル３０６や、さらに管理装置１６０のボリューム管理テーブル５０７や、連携サーバ１５０のボリューム管理テーブル４０５にも変更を行う。

【００２０】

次に本発明の第二の実施形態を図１４から図１６を参照して説明する。

第二の実施形態では、仮想ボリュームの構成に使用されていない比較的小さな容量の記憶領域が少なくなるように、既に仮想ボリュームを構成する一つ以上の記憶領域を別の記憶領域へと、運用中にデータ移行を行う場合について説明する。

図１４は、論理ユニット管理テーブルと仮想ボリュームテーブルの構成例を示す図であり、図７の構成例と同じである。

図１５はデータ移行を行う契機を得るための処理シーケンスの例を示す図である。

図１６はユニットの構成容易度を管理するユニット使用状況管理テーブル２１３の例を示す図である。

構成容易度は、一例として空きブロック数が７，５００未満の場合は「低」、１２，５００未満の場合は「中」、１２，５００以上の場合は「高」としている。

【００２１】

図１５の処理シーケンスにおいて、まず、ユニットの選択をする（ステップ１５０１）。例として、ユニットＲＵＡ０００１（図6、図7のエントリ６０４Ａ２）を選択したとする。

次に、記憶領域の構成容易度を調査し（ステップ１５０２）、ステップ１５０３に進む。

ステップ１５０３では、選択したユニットに構成容易度の低い領域があるか判定し、無ければデータ移行は行わず、有ればステップ１５０４に進む。例の場合、ユニットＲＵＡ０００１の各記憶領域の構成容易度を調査し、図１６のユニットＩＤがＲＵＡ００１のエントリには構成容易度のフィールド１６０５に「中」と「低」の記憶領域があることが判るので、図１０の処理シーケンスで示した処理ステップ１００４以降と同様の処理を、図１５の処理ステップ１５０４以降の処理で行う。

ただし、第二の実施形態では、移行元の論理ユニットのデータを最終的な移行先の論理ユニットへ直接移行できない場合に、少なくとも一つの間接的な論理ユニットに一旦データ移行し、その後に最終的な移行先の論理ユニットにデータ移行する場合を示している。

【００２２】

図１４において、（１）ユニットＲＵＡ０００Ｘに間接的な論理ユニットＶＬＵ０２の作成を行い、（２）ユニットＲＵＡ００００にある移行元の論理ユニットＶＬＵ０１からユニットＲＵＡ０００ＸにあるＶＬＵ０２にデータをコピーした後、ユニットＲＵＡ０００ＸにあるＶＬＵ０２のコピーしたデータを最終的な移行先であるユニットＲＵＡ００００にコピーし、このコピーしたデータに対して新たな論理ユニットＶＬＵ０１を設定し、（３）元の論理ユニットＶＬＵ０１を新たな論理ユニットＶＬＵ０１に切替え、（４）間接的な論理ユニットＶＬＵ０２の削除と（５）論理ユニットＶＬＵ０２内のデータ削除を行う。

図１６において、ユニットＲＵＡ０００Ｘは図１４の（１）でエントリ１６０７のように変更され、ユニットＲＵＡ０００１とＲＵＡ０００Ｘは図１４の（４）でエントリ１６０８のように変更され、最終的にユニットＲＵＡ０００１は構

成容易度が高くなっている。

【００２３】

次に本発明の第三の実施形態を図１７と１８を参照して説明する。

第三の実施形態では、スイッチの主要構成要素を冗長構成にした場合の説明と、管理装置からのテーブルの配布方法と、スイッチ管理部のテーブルの同期方法について説明する。

図１７はスイッチ１００の主要構成要素を冗長構成にした場合の計算機システムの構成例である。

図１７では、ルーティング制御部１２０、クロスバースイッチ１３０、スイッチ管理部１４０が冗長構成となっている。

第一の実施形態から第二の実施形態で示したような、ユニット管理テーブル６００、論理ユニット管理テーブル７０１、仮想ボリューム管理テーブル７０２は、図１８の情報の流れ１８０２のように管理装置１６０からスイッチ管理部１４０Ａと１４０Ｂに配布され、メインメモリ１４２Ａと１４２Ｂに格納される。

冗長構成のスイッチ管理部１４０は、現用系と予備系のように動作してもよいし、どちらも現用系で処理分担をしてもよい。

冗長構成のスイッチ管理部１４０の片系が障害となった場合は、現用系／予備系の場合は予備系に切替わり、どちらも現用系の場合は縮退運用となる。

冗長構成のスイッチ管理部１４０の持つテーブル１４３は、情報の流れ１８０１のように同期をとっておくことにより、片系が障害となった場合に比較的高速な切替えが可能となる。

【００２４】

次に本発明の第四の実施形態を図１９を参照して説明する。

第四の実施形態ではスイッチのスイッチ管理部がハードディスクを持つ構成の場合にデータ移行を行う場合について説明する。

図１９は、スイッチ１００のスイッチ管理部１４０がハードディスク１９０１を有する構成例を示した図である。

第一の実施形態から第二の実施形態で示したように、ハードディスク１９０１もユニット管理テーブル６００（図６参照）の一つとして扱うことにより、論理

ユニットの作成や仮想ボリュームへの構成を可能とする。

ユニット管理テーブル１９０３は、一例としてエントリ１９０４のような構成をとる。

ユニットＲＵＳ００００にデータ移行の際の間接的な論理ユニットを作成することにより、ストレージ装置１８０（図１参照）の記憶領域が比較的少なくなってきた場合に使用することが可能であり、データ移行時のデータがスイッチを通過する回数を低減することが可能となる。

【００２５】

【発明の効果】

本発明によれば、比較的少ない記憶装置からの記憶領域により仮想ボリュームを構成し、中継装置の中継処理負荷を低減させることができる。

また、比較的小さな容量の未使用の記憶領域の低減と、結果として比較的大きな容量の連続した空きの記憶領域の増加を図ることができる。

【図面の簡単な説明】

【図１】

計算機システムの構成例を示す図である。

【図２】

スイッチのスイッチ管理部が持つプログラムおよびテーブルの構成例を示す図である。

【図３】

スイッチのルーティング制御部が持つプログラムおよびテーブルの構成例を示す図である。

【図４】

スイッチの連携サーバが持つプログラムおよびテーブルの構成例を示す図である。

【図５】

計算機システムの管理装置が持つプログラムおよびテーブルの構成例を示す図である。

【図６】

ユニットの構成例とユニット管理テーブルの構成例を示す図である。

【図７】

論理ユニット管理テーブルと仮想ボリューム管理テーブルの構成例を示す図である。

【図８】

論理ユニット管理テーブルと仮想ボリューム管理テーブルの構成例を示す図である。

【図９】

データ移行の処理概要を示す図である。

【図１０】

データ移行の契機を得るための処理シーケンスの例を示す図である。

【図１１】

データ移行の処理シーケンスの例を示す図である。

【図１２】

論理ユニットの状態を管理するテーブルの例を示す図である。

【図１３】

データ移行中のＩ／Ｏ処理を行うためのブロックｂｉｔｍａｐテーブルとＩ／Ｏ処理マトリックスの例を示す図である。

【図１４】

データ移行の処理概要を示す図である。

【図１５】

ディスクの構成容易度をデータ移行の契機にする場合の処理シーケンスの例を示す図である。

【図１６】

ユニット使用状況管理テーブルの構成例を示した図である。

【図１７】

スイッチの主要構成要素を冗長構成とした場合の構成例を示した図である。

【図１８】

冗長構成のスイッチ管理部のテーブル情報の配布および同期の方法を示す図で

ある。

【図１９】

スイッチ管理部がハードディスクを持つ構成例を示す図である。

【符号の説明】

１００　スイッチ

１１０　接続インタフェース

１２０　ルーティング制御部

１３０　クロスバースイッチ

１４０　スイッチ管理部

１５０　連携サーバ

１６０　管理装置

１６１　ポート

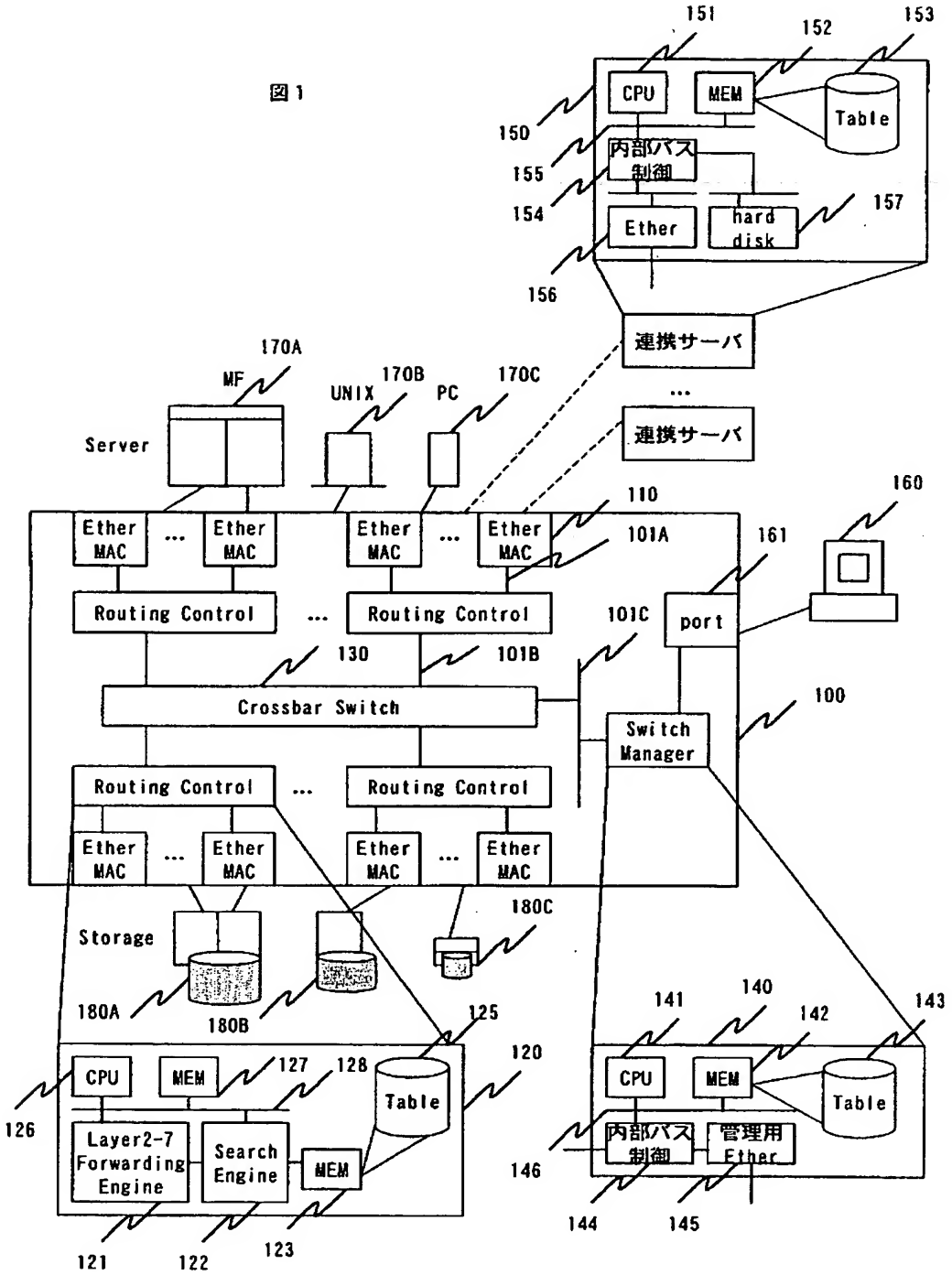
１７０　サーバ装置

１８０　ストレージ装置

【書類名】　図面

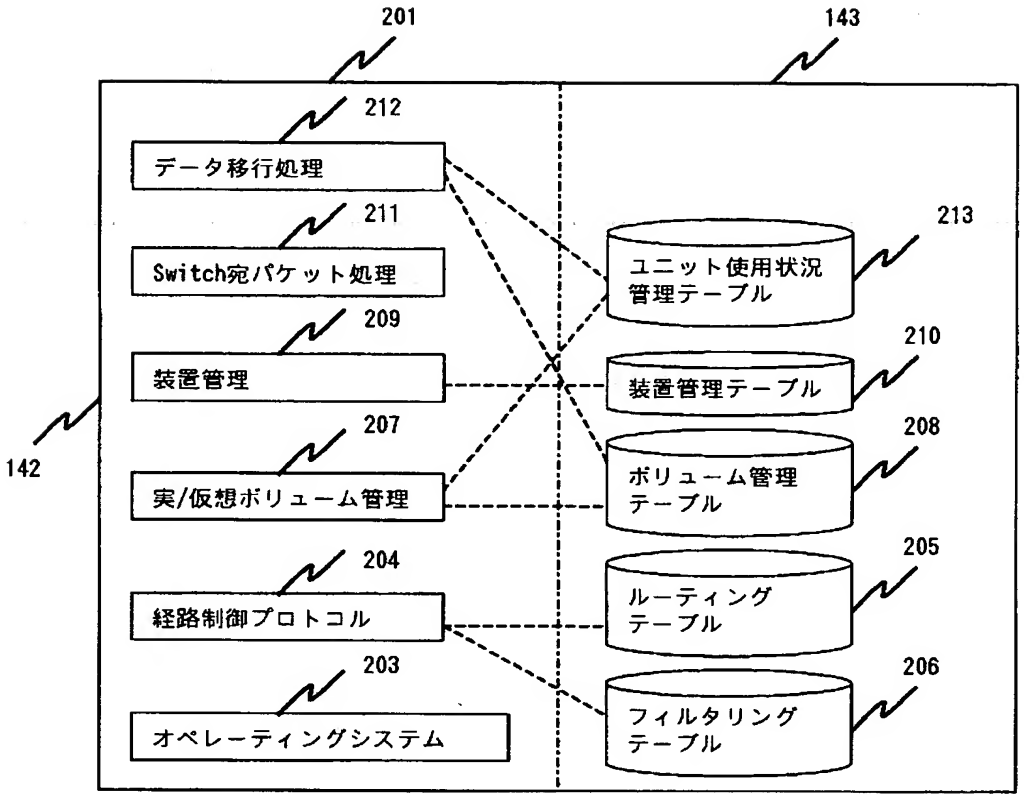
【図１】

図１


151
152
153
CPU
MEM
Table
150
155
内部バス制御
154
Ether
hard disk
157
156
連携サーバ
...
連携サーバ
170A
MF
170B
UNIX
170C
PC
Server
110
160
Ether MAC
101A
161
Routing Control
101C
port
130
101B
Crossbar Switch
100
Switch Manager
180C
Storage
180A
180B
125
141
140
142
143
120
CPU
MEM
127
128
Table
126
Layer2-7 Forwarding Engine
Search Engine
146
管理用 Ether
144
145
121
122
123

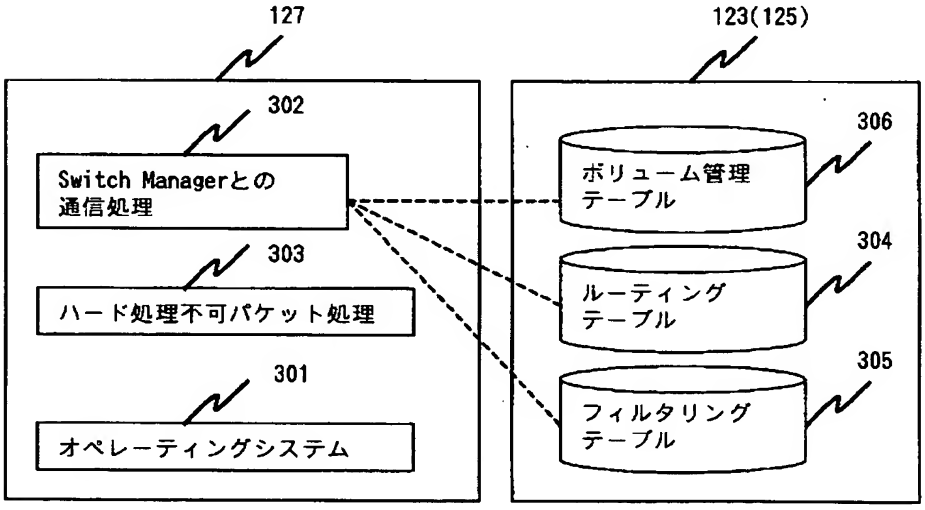
【図２】

図２


201
143
212
データ移行処理
211
Switch宛パケット処理
209
装置管理
207
実/仮想ボリューム管理
204
経路制御プロトコル
203
オペレーティングシステム
142
213
ユニット使用状況管理テーブル
210
装置管理テーブル
208
ボリューム管理テーブル
205
ルーティングテーブル
206
フィルタリングテーブル

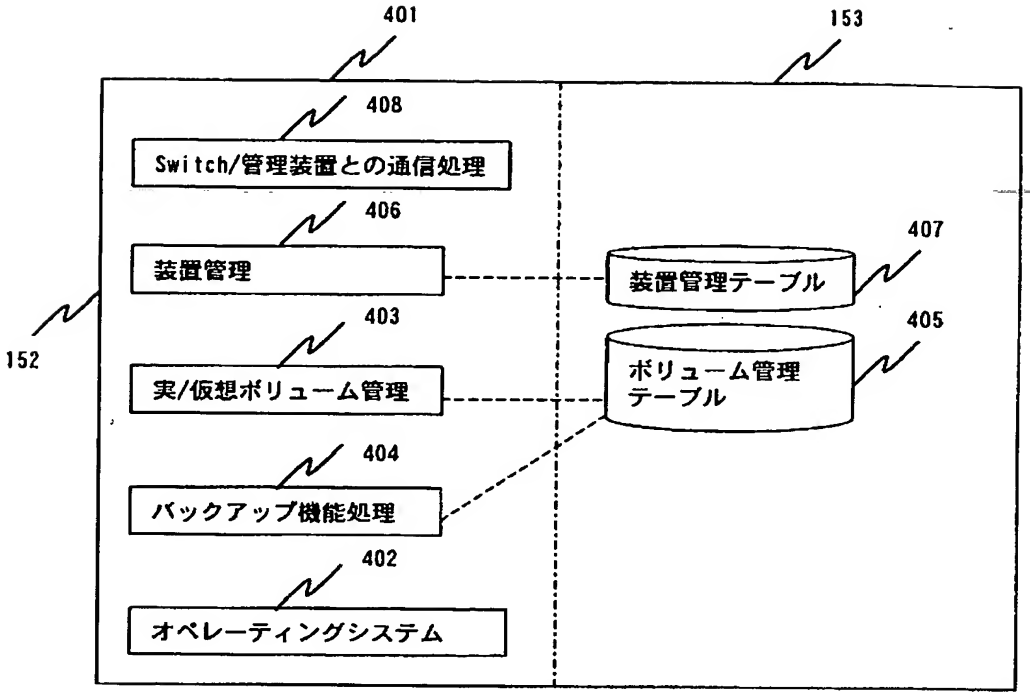
【図３】

図３


127
123(125)
302
Switch Managerとの通信処理
303
ハード処理不可パケット処理
301
オペレーティングシステム
306
ボリューム管理テーブル
304
ルーティングテーブル
305
フィルタリングテーブル

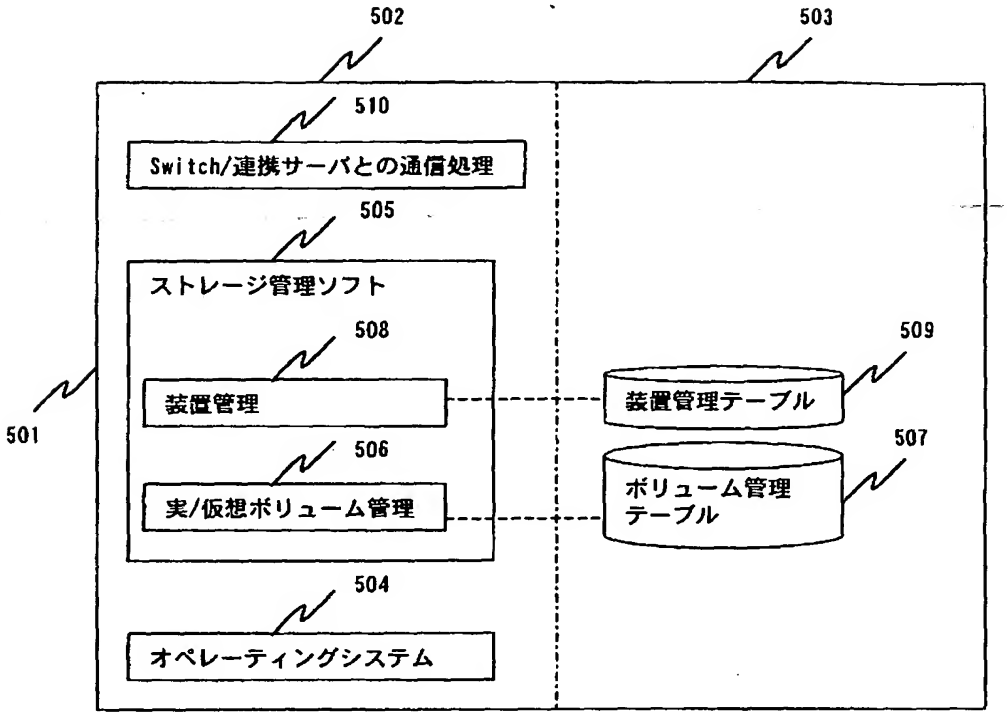
【図４】

図４


401
153
408
Switch/管理装置との通信処理
406
装置管理
407
装置管理テーブル
152
403
実/仮想ボリューム管理
405
ボリューム管理
テーブル
404
バックアップ機能処理
402
オペレーティングシステム

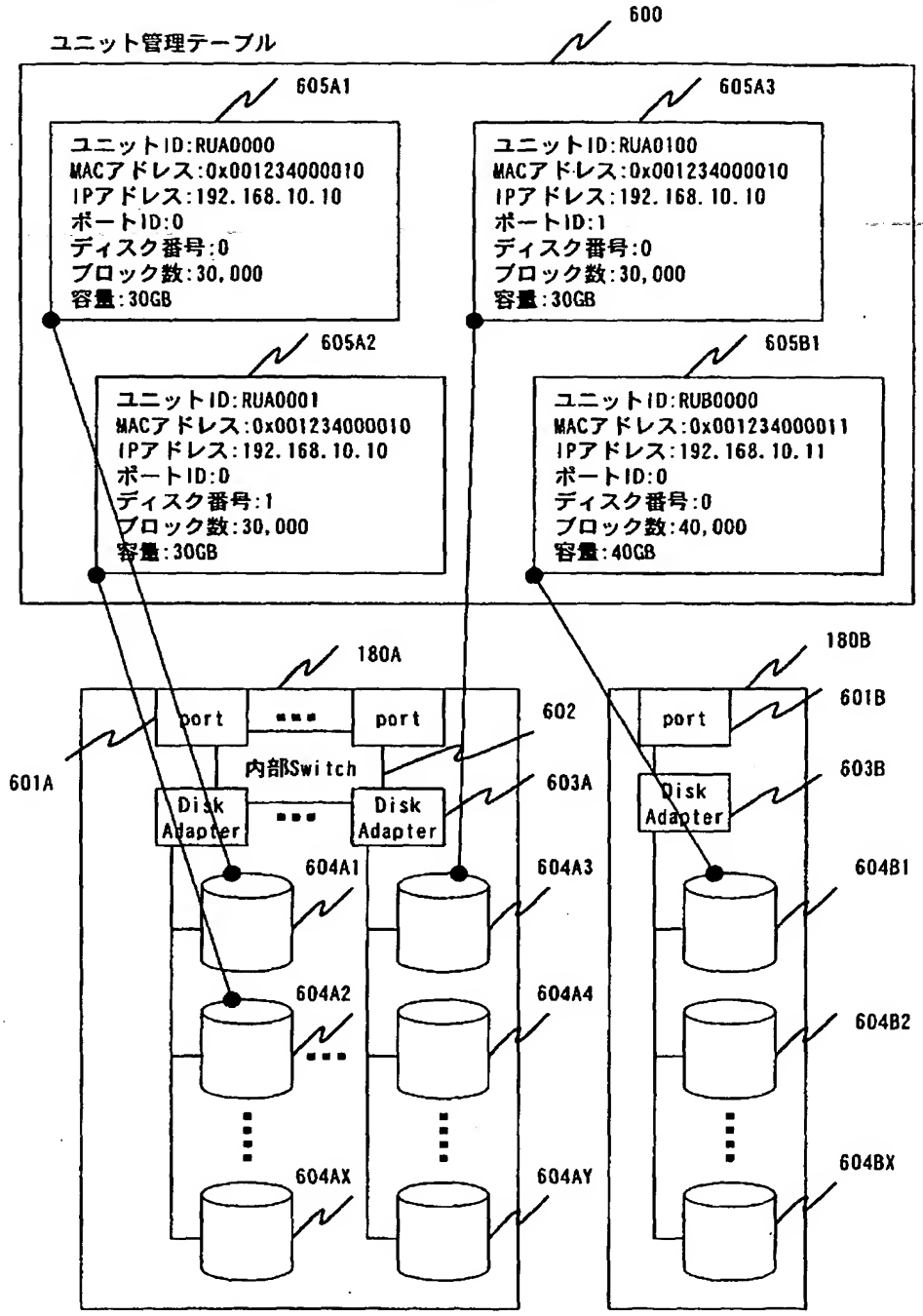
【図５】

図５


502
503
510
Switch/連携サーバとの通信処理
505
ストレージ管理ソフト
508
装置管理
509
装置管理テーブル
501
506
実/仮想ボリューム管理
507
ボリューム管理
テーブル
504
オペレーティングシステム

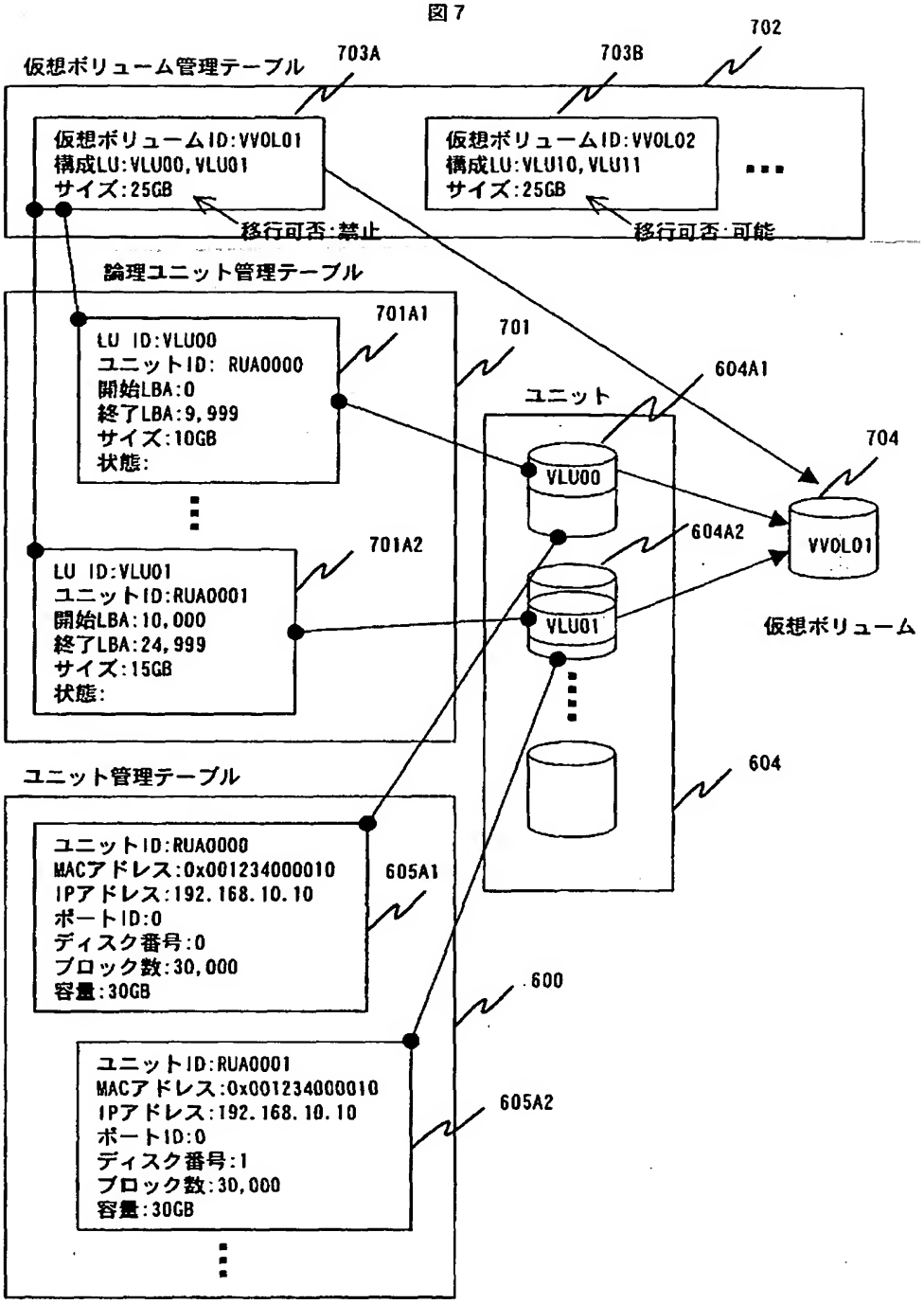
【図６】

図6


ユニット管理テーブル
600
605A1
ユニットID:RUA0000
MACアドレス:0x001234000010
IPアドレス:192.168.10.10
ポートID:0
ディスク番号:0
ブロック数:30,000
容量:30GB
605A3
ユニットID:RUA0100
MACアドレス:0x001234000010
IPアドレス:192.168.10.10
ポートID:1
ディスク番号:0
ブロック数:30,000
容量:30GB
605A2
ユニットID:RUA0001
MACアドレス:0x001234000010
IPアドレス:192.168.10.10
ポートID:0
ディスク番号:1
ブロック数:30,000
容量:30GB
605B1
ユニットID:RUB0000
MACアドレス:0x001234000011
IPアドレス:192.168.10.11
ポートID:0
ディスク番号:0
ブロック数:40,000
容量:40GB
180A
180B
port
port
port
601A
601B
602
内部Switch
603A
603B
Disk Adapter
Disk Adapter
Disk Adapter
604A1
604A3
604B1
604A2
604A4
604B2
604AX
604AY
604BX

【図７】

図７


仮想ボリューム管理テーブル
703A
703B
702
仮想ボリュームID:VVOL01
構成LU:VLU00, VLU01
サイズ:25GB
仮想ボリュームID:VVOL02
構成LU:VLU10, VLU11
サイズ:25GB
移行可否:禁止
移行可否:可能
論理ユニット管理テーブル
701A1
701
LU ID:VLU00
ユニットID: RUA0000
開始LBA:0
終了LBA:9,999
サイズ:10GB
状態:
LU ID:VLU01
ユニットID:RUA0001
開始LBA:10,000
終了LBA:24,999
サイズ:15GB
状態:
701A2
ユニット
604A1
604A2
VLU00
VLU01
704
VVOL01
仮想ボリューム
604
ユニット管理テーブル
ユニットID:RUA0000
MACアドレス:0x001234000010
IPアドレス:192.168.10.10
ポートID:0
ディスク番号:0
ブロック数:30,000
容量:30GB
605A1
600
ユニットID:RUA0001
MACアドレス:0x001234000010
IPアドレス:192.168.10.10
ポートID:0
ディスク番号:1
ブロック数:30,000
容量:30GB
605A2

【図８】

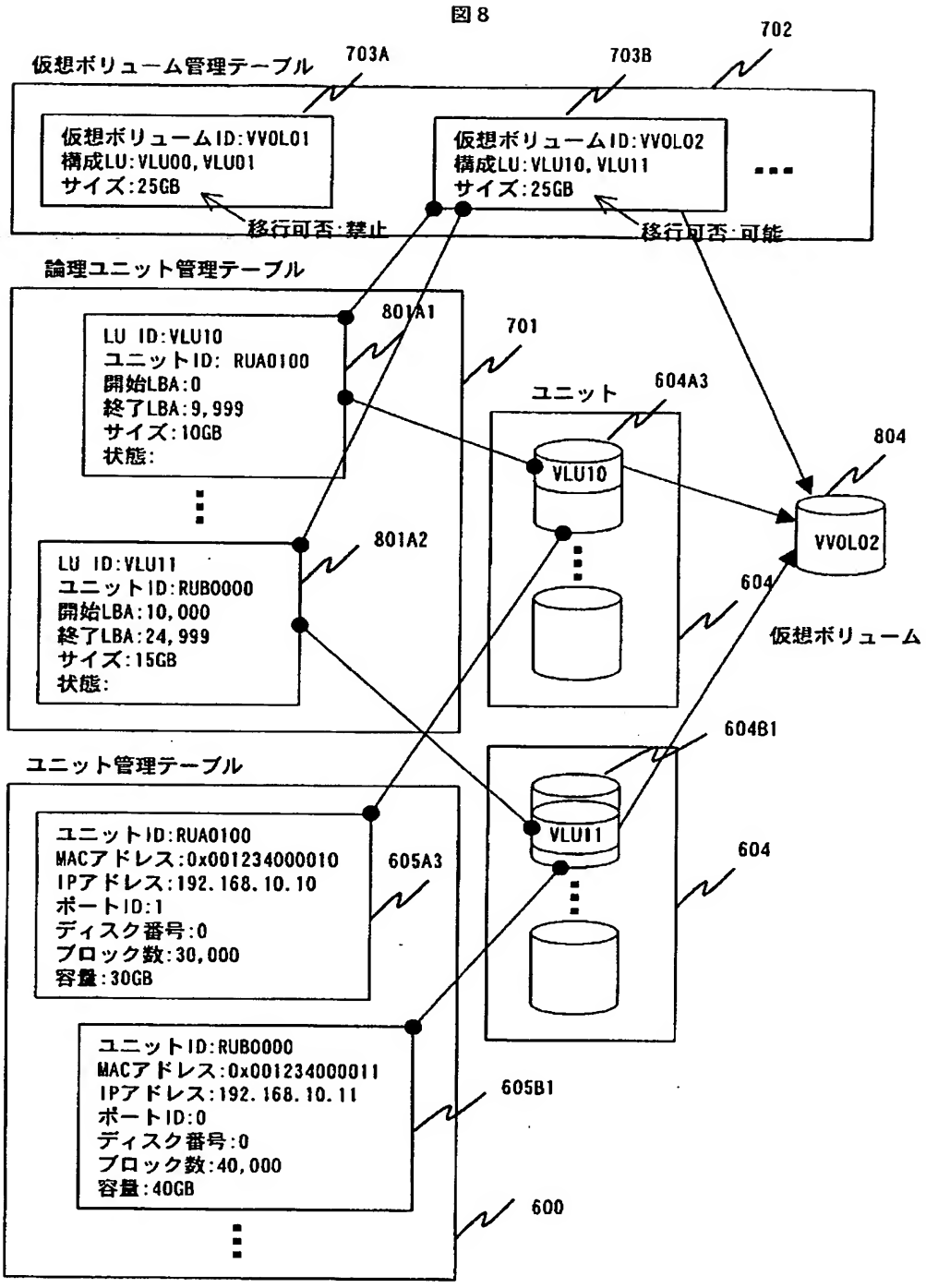
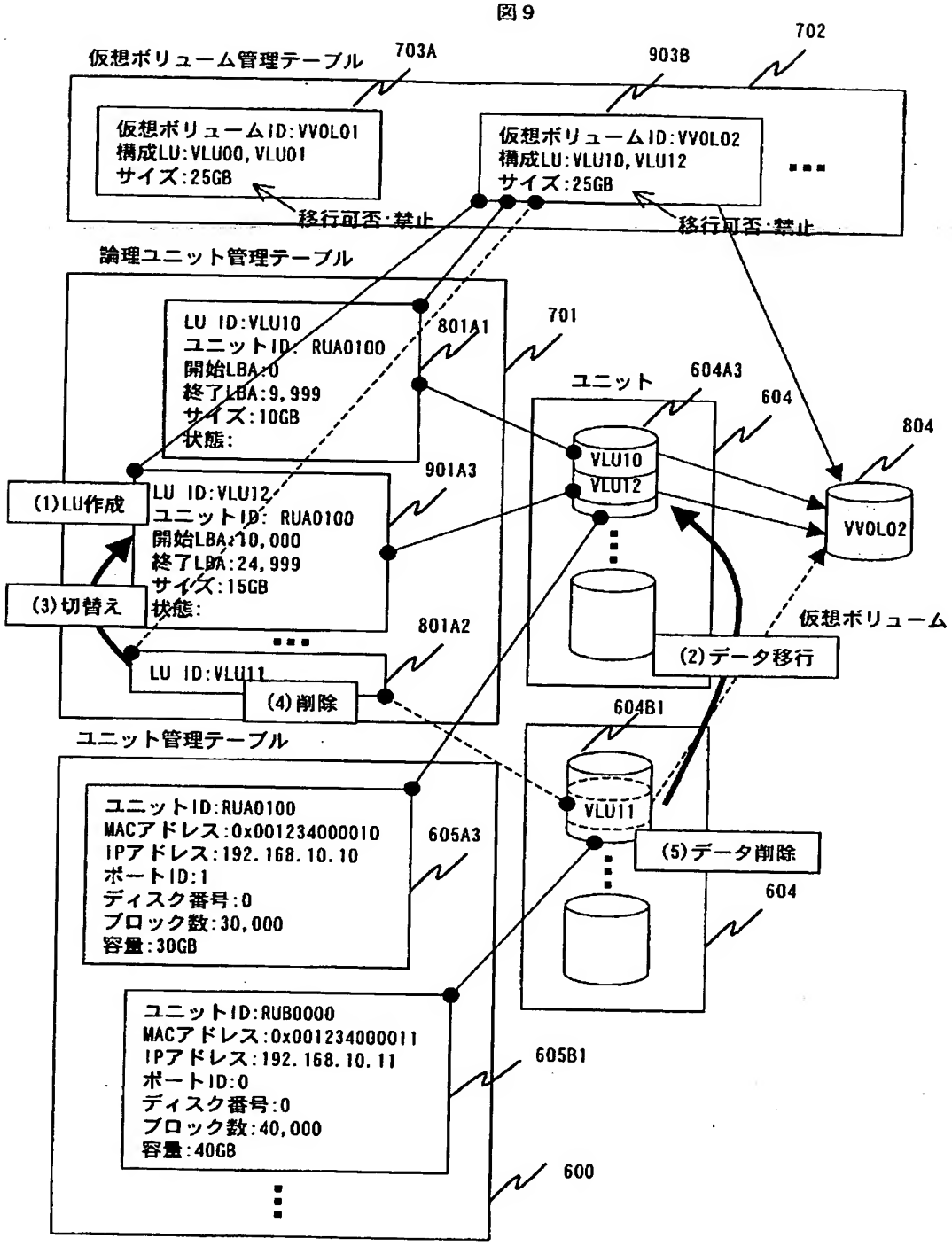


図８
仮想ボリューム管理テーブル
703A
703B
702
仮想ボリュームID:VVOL01
構成LU:VLU00, VLU01
サイズ:25GB
仮想ボリュームID:VVOL02
構成LU:VLU10, VLU11
サイズ:25GB
移行可否:禁止
移行可否:可能
論理ユニット管理テーブル
801A1
701
LU ID:VLU10
ユニットID: RUA0100
開始LBA:0
終了LBA:9,999
サイズ:10GB
状態:
801A2
LU ID:VLU11
ユニットID:RUB0000
開始LBA:10,000
終了LBA:24,999
サイズ:15GB
状態:
ユニット
604A3
VLU10
804
VVOL02
604
仮想ボリューム
604B1
VLU11
604
ユニット管理テーブル
ユニットID:RUA0100
MACアドレス:0x001234000010
IPアドレス:192.168.10.10
ポートID:1
ディスク番号:0
ブロック数:30,000
容量:30GB
605A3
ユニットID:RUB0000
MACアドレス:0x001234000011
IPアドレス:192.168.10.11
ポートID:0
ディスク番号:0
ブロック数:40,000
容量:40GB
605B1
600

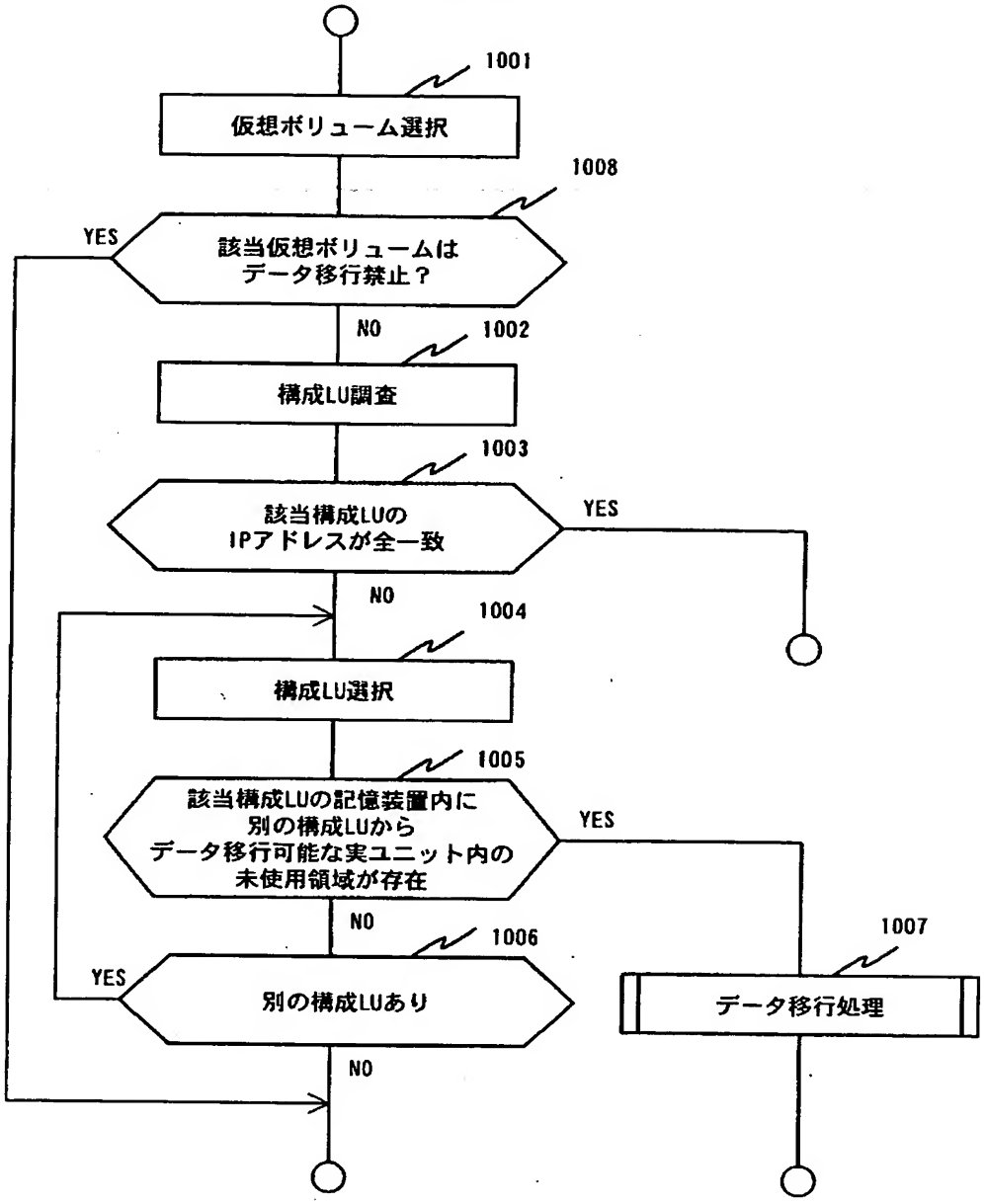
【図９】

図９


仮想ボリューム管理テーブル
703A
903B
702
仮想ボリュームID:VVOL01
構成LU:VLU00, VLU01
サイズ:25GB
仮想ボリュームID:VVOL02
構成LU:VLU10, VLU12
サイズ:25GB
移行可否:禁止
移行可否:禁止
論理ユニット管理テーブル
LU ID:VLU10
ユニットID: RUA0100
開始LBA:0
終了LBA:9,999
サイズ:10GB
状態:
801A1
701
ユニット
604A3
604
804
VLU10
VLU12
VVOL02
(1)LU作成
LU ID:VLU12
ユニットID: RUA0100
開始LBA:10,000
終了LBA:24,999
サイズ:15GB
状態:
901A3
(3)切替え
801A2
LU ID:VLU11
(4)削除
(2)データ移行
仮想ボリューム
604B1
ユニット管理テーブル
ユニットID:RUA0100
MACアドレス:0x001234000010
IPアドレス:192.168.10.10
ポートID:1
ディスク番号:0
ブロック数:30,000
容量:30GB
605A3
VLU11
(5)データ削除
604
ユニットID:RUB0000
MACアドレス:0x001234000011
IPアドレス:192.168.10.11
ポートID:0
ディスク番号:0
ブロック数:40,000
容量:40GB
605B1
600

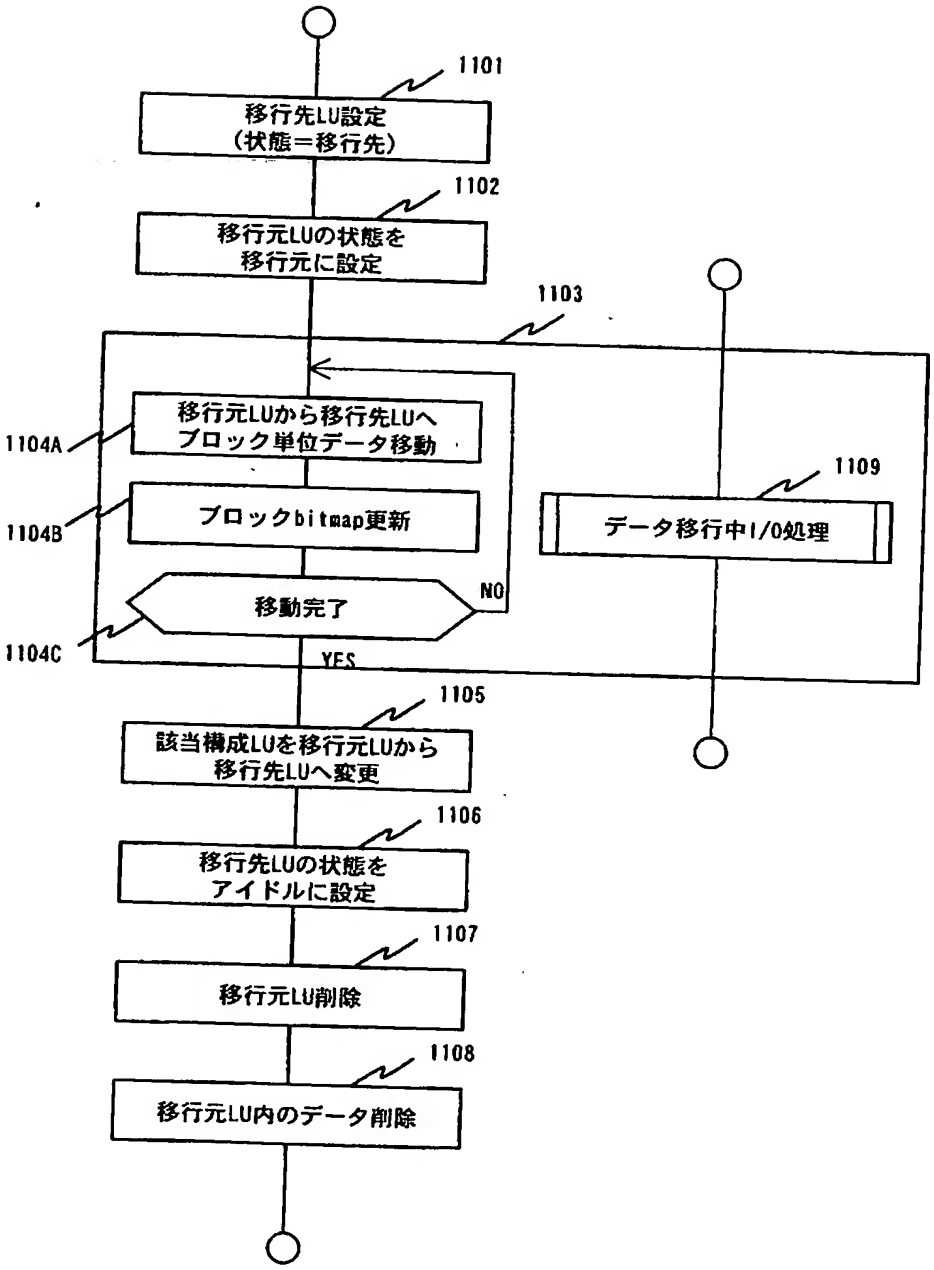
【図１０】

図１０


1001
仮想ボリューム選択
1008
該当仮想ボリュームは
データ移行禁止？
YES
NO
1002
構成LU調査
1003
該当構成LUの
IPアドレスが全一致
YES
NO
1004
構成LU選択
1005
該当構成LUの記憶装置内に
別の構成LUから
データ移行可能な実ユニット内の
未使用領域が存在
YES
NO
1006
別の構成LUあり
YES
NO
1007
データ移行処理

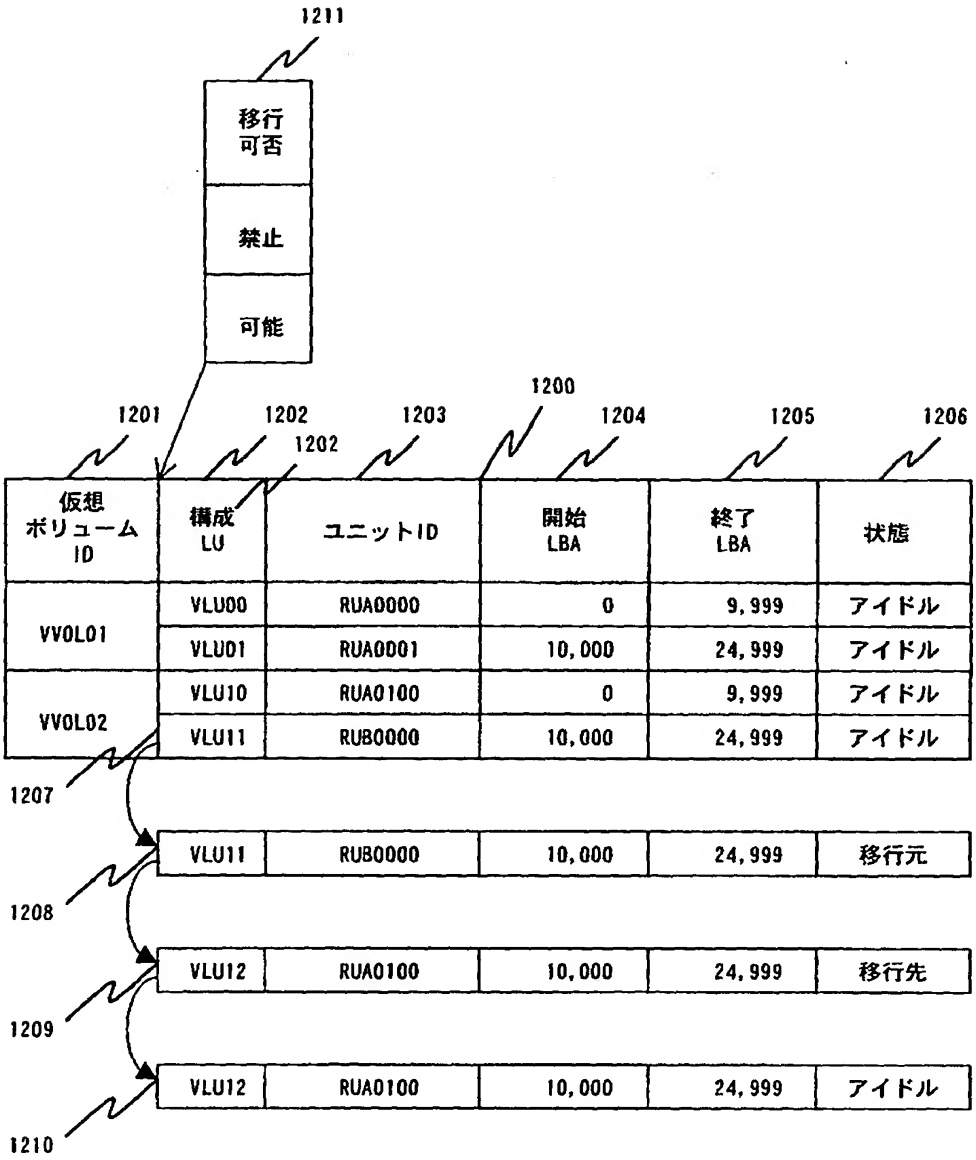
【図１１】

図１１


1101
移行先LU設定
（状態＝移行先）
1102
移行元LUの状態を
移行元に設定
1103
1104A
移行元LUから移行先LUへ
ブロック単位データ移動
1104B
ブロックbitmap更新
1104C
移動完了
NO
YES
1109
データ移行中I/O処理
1105
該当構成LUを移行元LUから
移行先LUへ変更
1106
移行先LUの状態を
アイドルに設定
1107
移行元LU削除
1108
移行元LU内のデータ削除

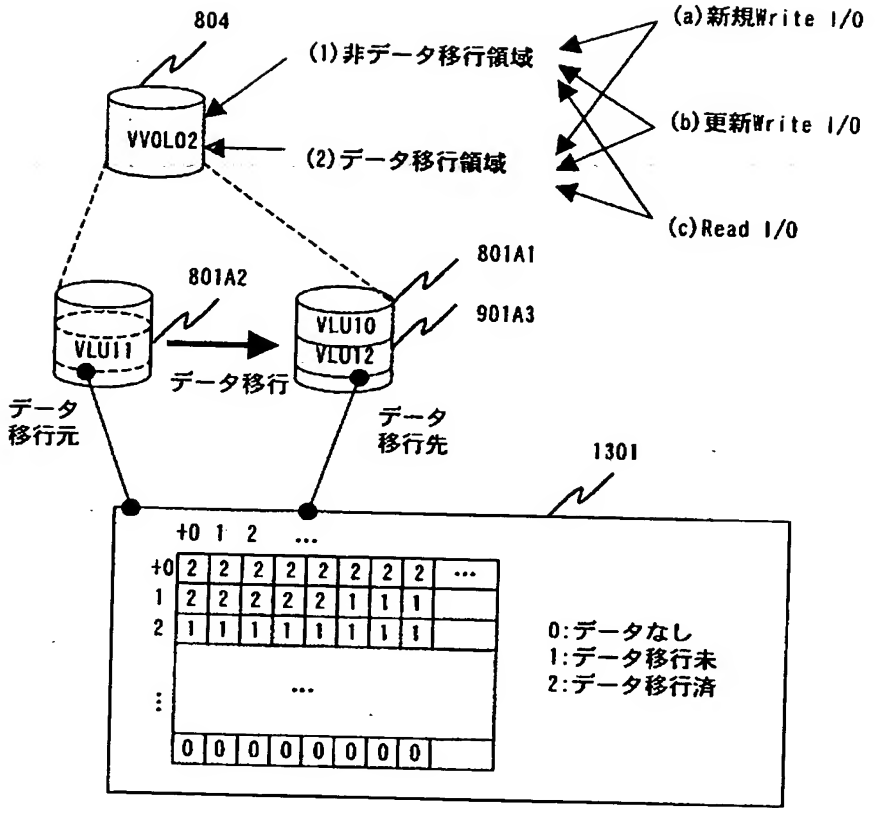
【図１２】

図１２


1211
移行可否
禁止
可能
1200
1201
1202
1202
1203
1204
1205
1206
仮想ボリュームID
構成LU
ユニットID
開始LBA
終了LBA
状態
VVOL01
VLU00
RUA0000
0
9,999
アイドル
VLU01
RUA0001
10,000
24,999
アイドル
VVOL02
VLU10
RUA0100
0
9,999
アイドル
VLU11
RUB0000
10,000
24,999
アイドル
1207
VLU11
RUB0000
10,000
24,999
移行元
1208
VLU12
RUA0100
10,000
24,999
移行先
1209
VLU12
RUA0100
10,000
24,999
アイドル
1210

【図１３】

図１３

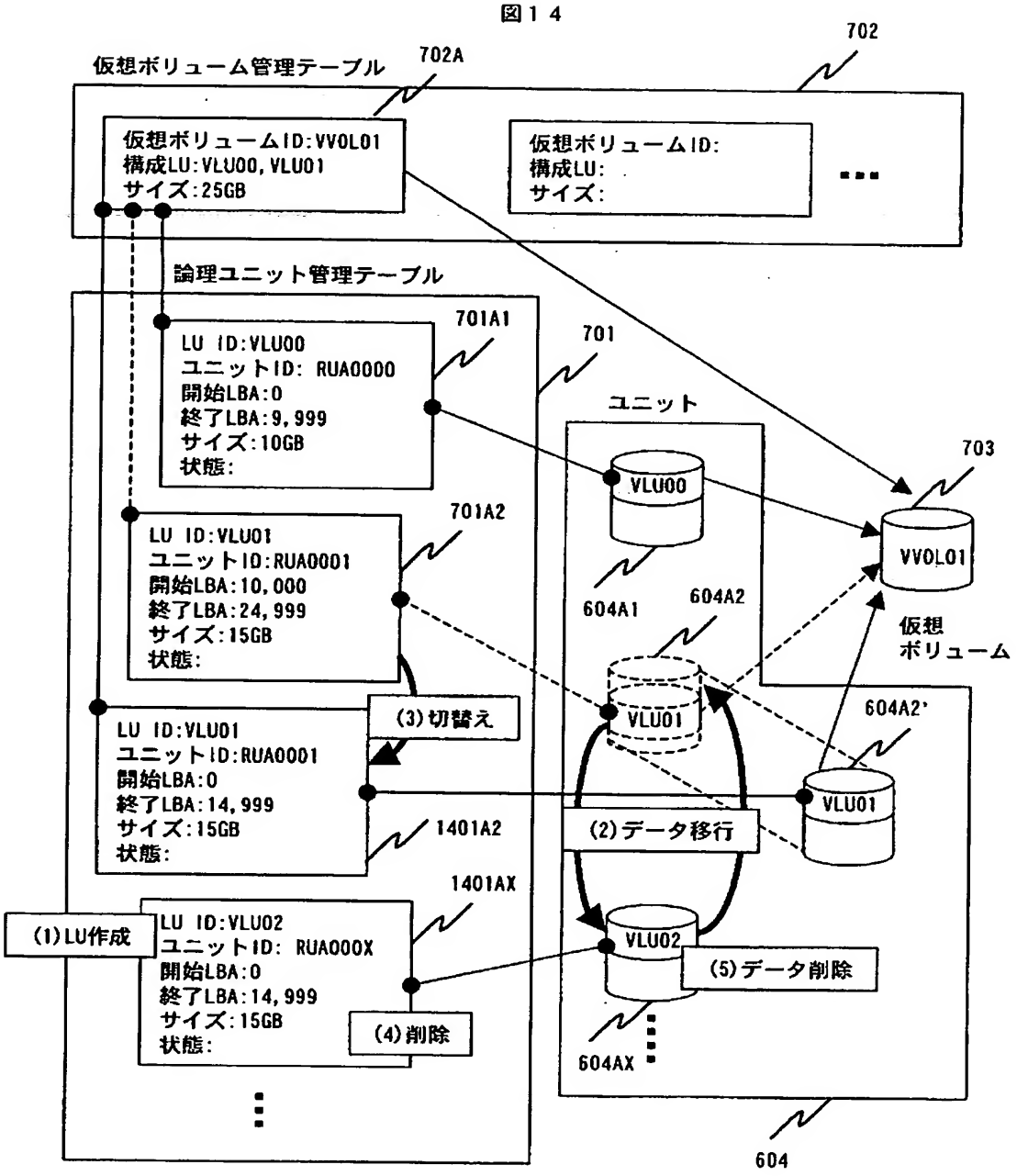


1302

アクセス先 / 入力I/O	(1)非データ移行領域	(2)データ移行領域	
		データ移行未（該当ブロックbitmap=1）	データ移行済（該当ブロックbitmap=2）
(a)新規Write	該当領域へ	データ移行先のブロックbitmap=0領域へ、Write後bitmap=2	
(b)更新Write	該当領域へ	・データ移行元へor ・データ移行先へ、該当ブロックbitmap=2	データ移行先へ
(c)Read	該当領域へ	データ移行元へ	データ移行先へ

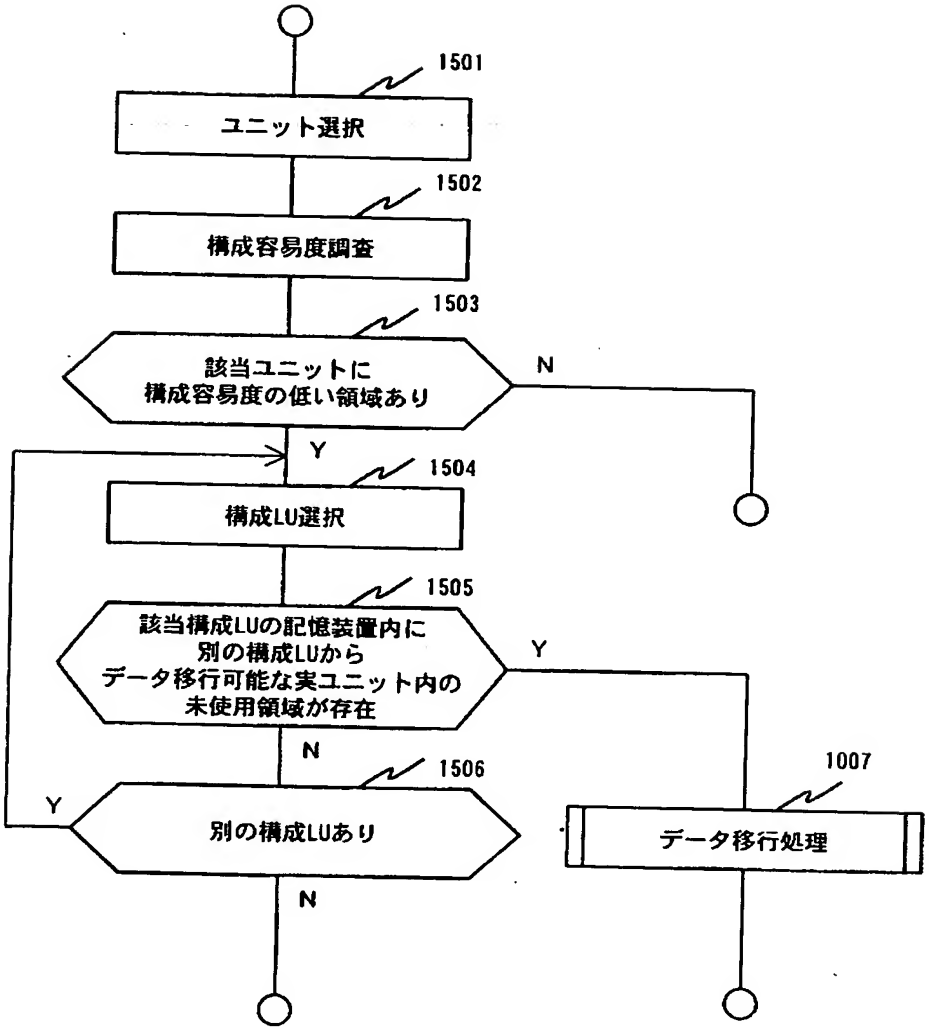
【図１４】

図１４


仮想ボリューム管理テーブル
702A
702
仮想ボリュームID:VVOL01
構成LU:VLU00, VLU01
サイズ:25GB
仮想ボリュームID:
構成LU:
サイズ:
論理ユニット管理テーブル
701A1
701
LU ID:VLU00
ユニットID: RUA0000
開始LBA:0
終了LBA:9,999
サイズ:10GB
状態:
ユニット
703
VLU00
VVOL01
701A2
LU ID:VLU01
ユニットID:RUA0001
開始LBA:10,000
終了LBA:24,999
サイズ:15GB
状態:
604A1
604A2
仮想
ボリューム
(3)切替え
VLU01
604A2'
LU ID:VLU01
ユニットID:RUA0001
開始LBA:0
終了LBA:14,999
サイズ:15GB
状態:
1401A2
VLU01
(2)データ移行
1401AX
(1)LU作成
LU ID:VLU02
ユニットID: RUA000X
開始LBA:0
終了LBA:14,999
サイズ:15GB
状態:
VLU02
(5)データ削除
(4)削除
604AX
604

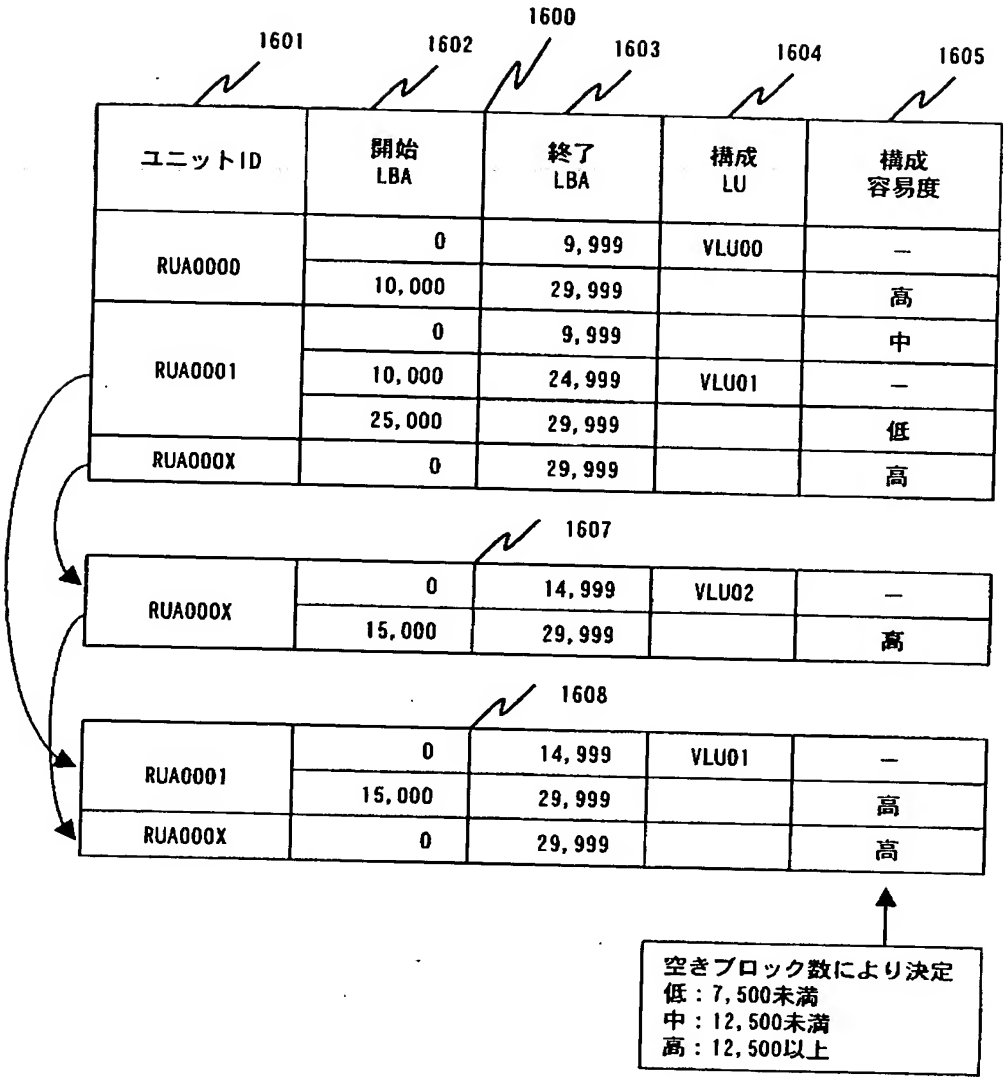
【図１５】

図１５


1501
ユニット選択
1502
構成容易度調査
1503
該当ユニットに
構成容易度の低い領域あり
N
Y
1504
構成LU選択
1505
該当構成LUの記憶装置内に
別の構成LUから
データ移行可能な実ユニット内の
未使用領域が存在
Y
N
1506
Y
別の構成LUあり
N
1007
データ移行処理

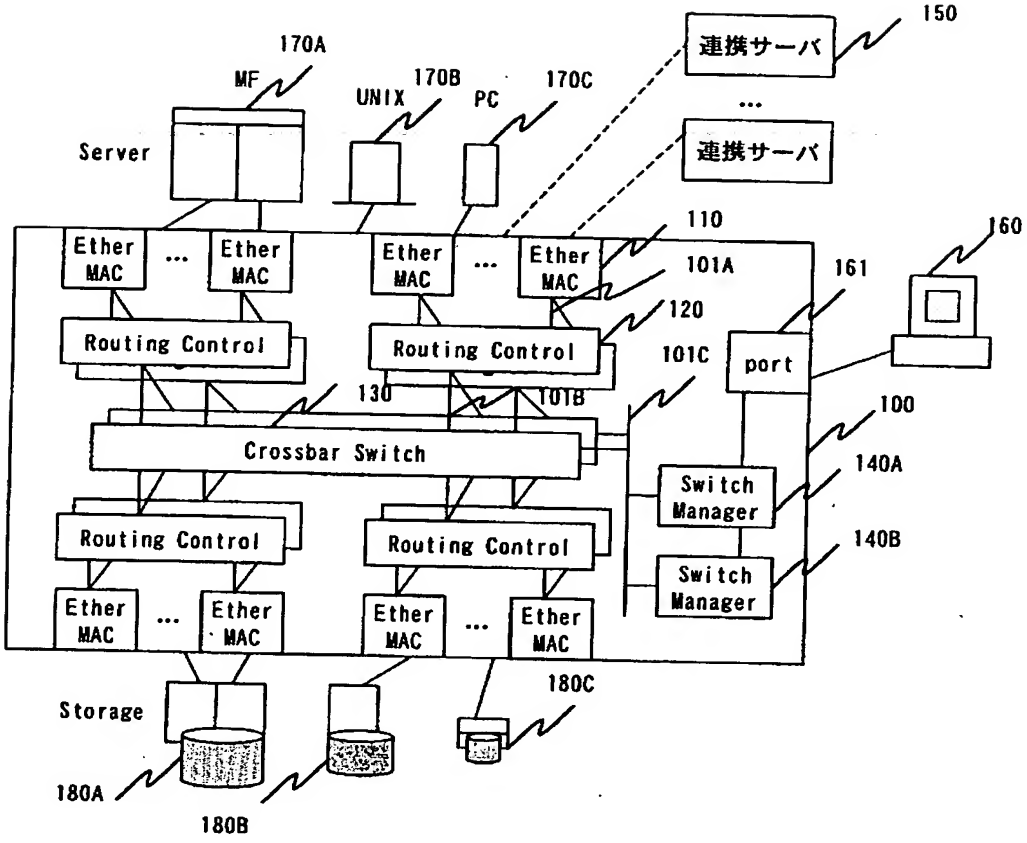
【図１６】

図１６


1600
1601
1602
1603
1604
1605
ユニットID
開始 LBA
終了 LBA
構成 LU
構成 容易度
RUA0000
0
9,999
VLU00
—
10,000
29,999
高
RUA0001
0
9,999
中
10,000
24,999
VLU01
—
25,000
29,999
低
RUA000X
0
29,999
高
1607
RUA000X
0
14,999
VLU02
—
15,000
29,999
高
1608
RUA0001
0
14,999
VLU01
—
15,000
29,999
高
RUA000X
0
29,999
高
空きブロック数により決定
低：7,500未満
中：12,500未満
高：12,500以上

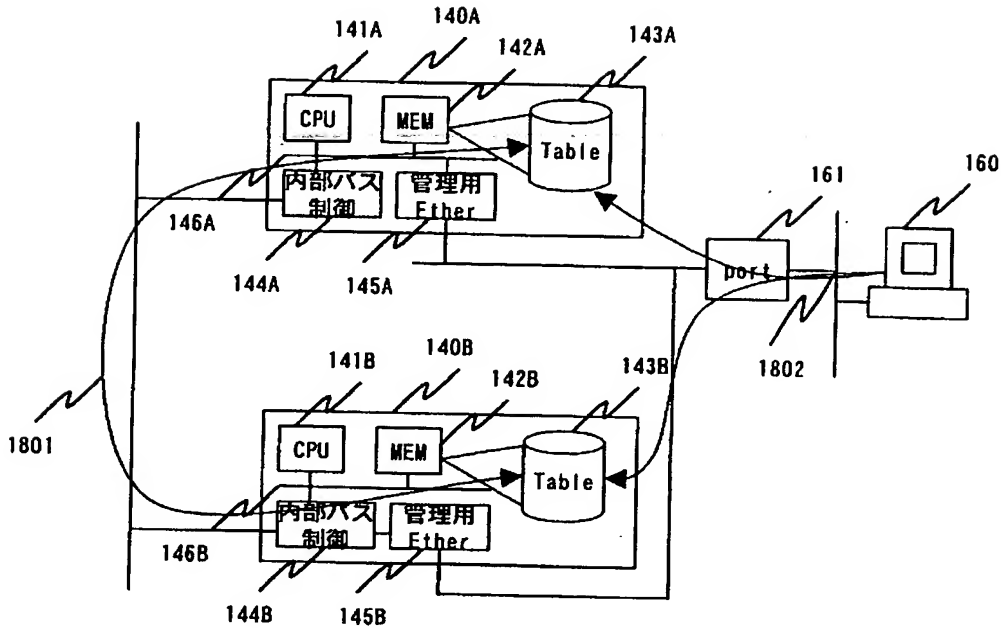
【図１７】

図１７


150
連携サーバ
...
連携サーバ
170A
MF
170B
UNIX
170C
PC
Server
110
160
Ether MAC
...
Ether MAC
Ether MAC
...
Ether MAC
101A
161
120
Routing Control
Routing Control
101C
port
130
101B
100
Crossbar Switch
140A
Switch Manager
Routing Control
Routing Control
140B
Switch Manager
Ether MAC
...
Ether MAC
Ether MAC
...
Ether MAC
180C
Storage
180A
180B

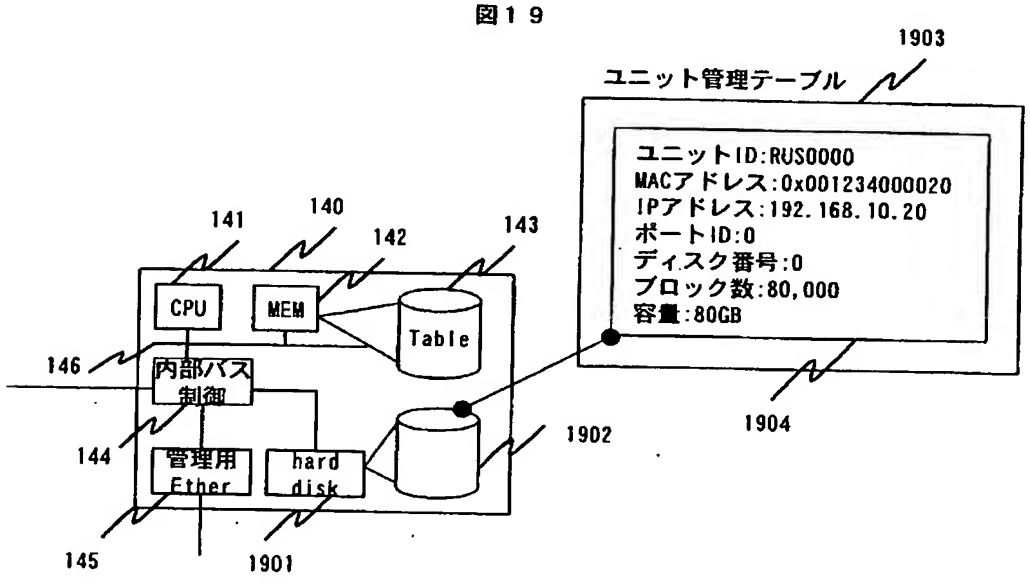
【図１８】

図１８



【図１９】

図１９



【書類名】　　要約書

【要約】

【課題】　比較的少ない数の記憶装置からの記憶領域により仮想ボリュームを構成するように、中継装置が運用中にデータ移行を行うことにより中継処理負荷を低減させる方法を提供する。

【解決手段】　ある一つの仮想ボリュームを構成している記憶装置（180）の宛先が増加してくると、中継装置(100)がこれを契機にできるだけ少ない記憶装置の宛先になるように仮想ボリュームの構成を変更できるように、複数の記憶装置間でデータ移行を行う。また、仮想ボリュームを構成していない未使用の記憶領域の中で比較的小さな容量の記憶領域が増加してくると、中継装置がこれを契機にできるだけこの小さな容量の記憶領域の数を低減するように、複数の記憶装置間でデータ移行を行う。

【選択図】　　図１

特願２００３－０３８０９７

出　願　人　履　歴　情　報

識別番号　　［０００００５１０８］

１．変更年月日　　１９９０年　８月３１日
［変更理由］　　新規登録
住　所　　東京都千代田区神田駿河台４丁目６番地
氏　名　　株式会社日立製作所